A Stereo Matching Algorithm for Urban Digital Elevation Models

C. Baillard and O. Dissard

Abstract

A stereo matching algorithm dedicated to complex urban scenes is described. It relies on successive complementary matching steps, based on dynamic programming. First, intensity edges of both images are matched, which produces piecewise continuous 3D chains. This provides a description of the scene structure containing the highest elevation of most height discontinuities. Then the interval pairs defined by the matched edges are matched in a hierarchical way, by a radiometrically constrained process followed by a geometrically constrained one.

The novelty of the approach lies in the use of several successive steps appropriate to different kinds of pixels. It provides dense disparity maps with less noise, while preserving discontinuities, which are a characteristic of urban digital elevation models. The method has proved reliable (producing few noisy and altimetrically accurate 3D data) and fast, and is robust to image variability. Perspectives within an industrial production context are discussed.

Introduction

The production of digital elevation models (DEM) from images has been one of the primary goals of cartography for many years. Recently, the area has been stimulated by the need for urban DEMs within many fields such as the production and management of three-dimensional (3D) databases, for urban and town planning, and for the simulation of line-of-sight microwave propagation. These applications usually require dense and reliable 3D data which preserve the characteristics of the scene. Particular attention has to be paid to height discontinuities, which are very frequent in urban environments.

Because of the complexity and the rapid development of urban scenes, producing and updating an urban DEM requires much time from a human operator. Therefore, a lot of work has been done to automatically derive a DEM from aerial stereo imagery. Figure 1 shows an example of an aerial stereo pair. The stereoscopic matching process is very difficult in urban environments. The high density of above-ground objects leads to many hidden parts or occlusions in the images, and, therefore, to many pixels without a correspondence in the other image. The difficulty is compounded by the fact that many scene features are similar to each other (parallel borders of buildings or roads), homogeneous areas are frequent (shadows, roofs), and moving objects (car, trucks) may disrupt the scene. These characteristics of urban scenes make the matching process very ambiguous. In addition, large depth discontinuies are common; therefore, geometric constraints about the surface must be used very carefully. Many matching algorithms have been proposed in the literature, but none of them have achieved satisfactory results in complex urban areas.

In this paper, a new matching algorithm especially dedicated to urban scenes is proposed. The input data are two stereo images sampled in epipolar geometry. The novelty of the approach lies in the use of several successive steps appropriate to different kinds of pixels. The algorithm provides dense 3D data, while preserving discontinuies.

The paper is organized as follows: A brief review of stereo reconstruction methods in an urban environment is first presented. The approach proposed in this paper is then introduced. It consists of two stages. First, intensity edges from both images are matched. Then dense elevation data are produced using an area-based matching process. The results are presented and discussed and the application of the technique to the industrial production of a DEM is presented.

Related Work

Much work has been done on automatic stereoscopic matching, and two distinct matching methods have emerged: feature-based and area-based approaches.

Feature-based matching consists of matching primitive sets extracted from each image. Common features in an urban environment are points of interest, segments, and linear structures (Horaud and Skordas, 1989; Hoff and Ahuja, 1989; Roux and McKeown, 1994; Noronha and Nevatia, 1997; Schmid and Zisserman, 1997). The feature-based approach is appropriate for discontinuity management, because height discontinuities often appear as intensity discontinuities in the images. However, an interpolation or an approximation of the surface is required to derive a dense DEM from the 3D features. In an urban environment, a common hypothesis consists of looking for planar surfaces. Thus, 3D linear features and planar approximations can be used for the 3D reconstruction of some buildings. However, the complexity of the surface in urban areas, which includes vegetation, arbitrarily complex buildings, and house aggregates, usually does not allow such an interpolation. The use of multiple views and very high resolution imagery (15 cm per pixel or less) appears to be necessary (Haala and Hahn, 1995; Bignone *et al.*, 1996; Moons *et al.*, 1998; Henricsson, 1998; Baillard *et al.*, 1999).

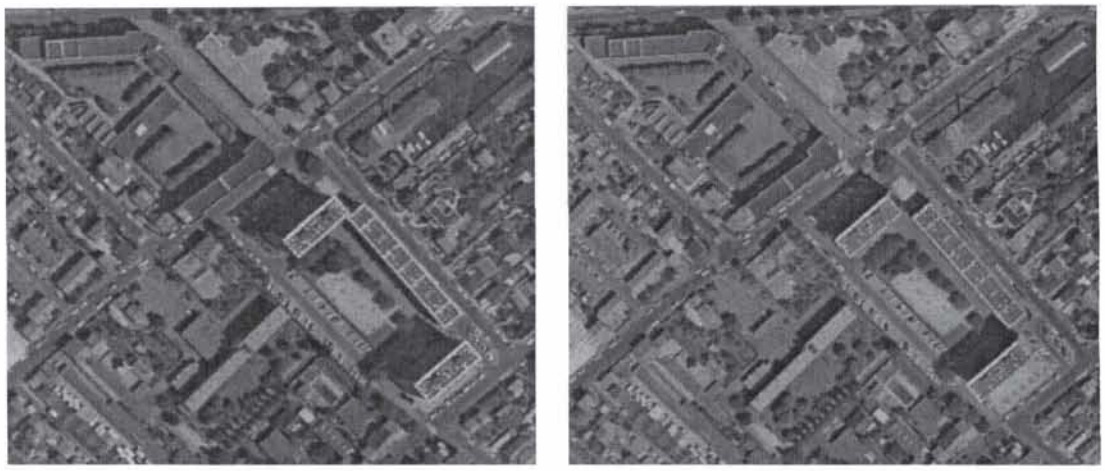
The second class of methods is area-based and aims at matching every pixel by measuring the similarity of grey-level templates. The advantage is to produce dense disparity maps and therefore dense DEMs. However, it is difficult to cope with both homogeneous areas and discontinuities. Homogeneous

Institut Géographique National, Laboratoire MATIS, 2 av. Pasteur, 94160 Saint Mandé, France

C. Baillard is currently with SIRADEL, Espace performance 3, bat M1, 35769 Saint-Gregoire, France (cbaillar@irisa.fr).

Photogrammetric Engineering & Remote Sensing
Vol. 66, No. 9, September 2000, pp. 1119–1128.

0099-1112/00/6609–1119$3.00/0



(a) (b)

Figure 1. Example of an aerial stereo pair, sampled in epipolar geometry. The images are 710 by 600 pixels, one pixel corresponding to a ground distance of 40 cm. A one-pixel difference in disparity (measured between the two images) corresponds to a 96-cm difference in height. The disparity range is about of 50 pixels.

areas are composed of ambiguous pixels for the matching process, and they introduce many errors in the DEM. Applying a continuity constraint between two neighboring 3D points is very efficient at reducing errors, and has been successfully used for the reconstruction of continuous surfaces or for low resolution imagery (typically satellite imagery) (Otto and Chau, 1989). Multi-resolution approaches are also frequently used to reduce errors in homogeneous areas: i.e., images are sub-sampled at various resolutions, and the matching process at a given resolution level is driven by disparity values produced at a coarser level (Hannah, 1989; Dhond and Aggarwal, 1995). This kind of approach also implicitely relies on a continuity assumption. These strategies must therefore be used very carefully when dealing with discontinuous surfaces (O'Neill and Denos, 1992). Another solution consists of involving more than two images in the matching process, because ambiguities and hidden part effects are then reduced (Berthod *et al.*, 1995; Leloglu *et al.*, 1998).

A common approach consists in combining an area-based matching process with a monocular analysis. It has been proposed to involve correlation windows with a variable size derived from the intensity edges (Lotti and Giraudon, 1994). More recently, an adaptive window shape based on intensity edges has also been proposed (Cord *et al.*, 1998). The surface can be *a posteriori* refined by adaptive correlation windows depending on the intensity function (Kanade and Okutomi, 1994), or by an adaptive smoothing (Cochran and Medioni, 1992). In Fua (1991), surfaces are interpolated by a diffusion algorithm driven by the radiometric gradient. Radiometric segmentations can be utilized to perform a planar approximation of 3D points with homogeneous radiometry (Maitre and Luo, 1992; Girard *et al.*, 1998; Fradkin *et al.*, 1999). Unfortunately, these methods suffer from segmentation limits. Many height discontinuities do not appear in the intensity image, and, reciprocally, many textural radiometric edges create over-segmentation problems. In order to overcome these limits, specific models of buildings must be used for approximating the DEM (Dang *et al.*, 1994; Weidner and Foerstner, 1995; Haala and Hahn, 1995; Paparoditis *et al.*, 1998). However, these models are not appropriate for dense urban scenes, where objects are complex and adjacent to each other.

An alternative solution to cope with discontinuities consists of explicitely involving occlusions in the matching process (Belhumeur and Mumford, 1992; Geiger *et al.*, 1995; Luo and Burkhardt, 1995; Birchfield and Tomasi, 1998). Occlusions are very important because a height discontinuity in the scene will generally produce a depth discontinuity in one view and an occlusion area in the other view.

Overview of the Matching Algorithm

The method presented in this paper takes advantage of both feature-based and area-based approaches, and takes occlusions into account through a dynamic programming optimization process. The novelty of the approach lies in the use of several successive steps, each of them appropriate for a certain kind of pixels. The input data are two stereo images sampled in an epipolar geometry (see Figure 1). The overview of the algorithm is presented in Figure 2.

The first stage consists in matching intensity edges from both images, which produces piecewise continuous 3D chains (see the section on Edge Point Matching). This provides a description of the scene structure containing the highest elevation of most height discontinuities.

The second stage is based on area-based matching (see the section on Area-Based Matching between Edge Points). It aims at matching the pairs of intervals defined by the matched edges from the first stage. Involving matched edges as anchor points reduces the search space for corresponding pixels, therefore decreasing the number of false matches and the computation time. A matched edge pair can be rejected if it is not consistent with the new matching set. This area-based matching stage itself works in two steps. First, a strong radiometric similarity constraint is applied, in order to produce only reliable pairs. Then a second matching step is performed with a looser radiometric constraint but a stronger geometric one (smoothness constraint), in order to complete the matching on unmatched areas. This hierarchical strategy relies on the assumption that local extrema of depth along epipolar lines are recovered as reliable pairs during the radiometrically constrained step. Some information about image radiometry and shadows is also computed and exploited during the matching. These two steps are described in later.

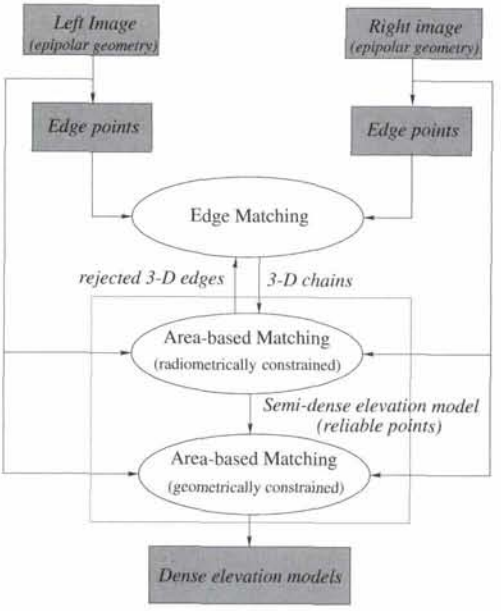


Figure 2. Overview of the matching process: A hierarchical strategy, consisting of a feature-based matching stage followed by a two-step area-based matching stage.

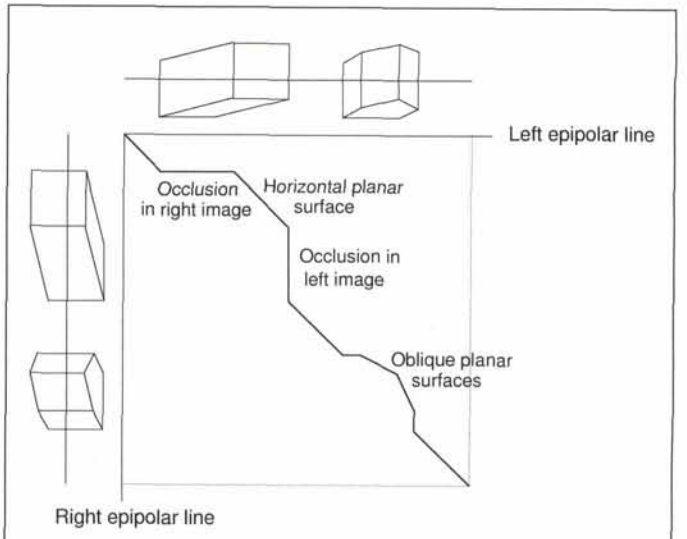


Figure 3. Correspondence graph and matching path. The axes of the graph represent conjugate epipolar lines. The subparts of the path which are parallel to the axes correspond to hidden areas in the right or the left image; the other subparts of the graph describe the visible 3D surface of the scene.

Each matching step is performed by dynamic programming. This is an efficient technique to solve constrained non-linear optimization problems, which is often used in stereovision to match edges or intervals. In this context, the matching problem is expressed as an optimization problem for each epipolar line pair. It consists of finding an optimal path in a 2D graph defined by conjugate epipolar lines, called the correspondence graph (see Figure 3). Dynamic programming is a powerful matching strategy, because it provides an optimal solution for each epipolar line pair, involving all consistency constraints along these lines: unicity, order, but also duality between discontinuity and occlusion. In particular, the hidden parts are not matched. Therefore, local constraints such as disparity range or thresholds on similarity measure can be released without significantly increasing errors. This aspect is very important for urban scenes where differences in height can be very large in the neighborhood of towers, and the correlation values can be very low in homogeneous areas.

The optimal path has to minimize a cost function $C(i,j)$ which is defined recursively within the disparity range by the relations

$$C(i,j) = \min \begin{cases} C(i-1, j-1) + match(i, j) & (1) \\ C(i-1, j) + occ(i) & (2) \\ C(i, j-1) + occ(j) & (3) \end{cases}$$

where *match* and *occ* are elementary cost functions associated with elementary paths according to the scheme of Figure 4. The elementary path (1) denotes a match between two edge points, and it is associated with the matching cost $match(i,j)$. The elementary paths (2) and (3) correspond to an occluded point in the right image or the left image, and they are both associated with the cost function *occ*. Such a configuration is symmetrical to both images. If the disparity value related to the pair (L_i, R_j) is not a member of the disparity range, then $C(i,j) = +\infty$.

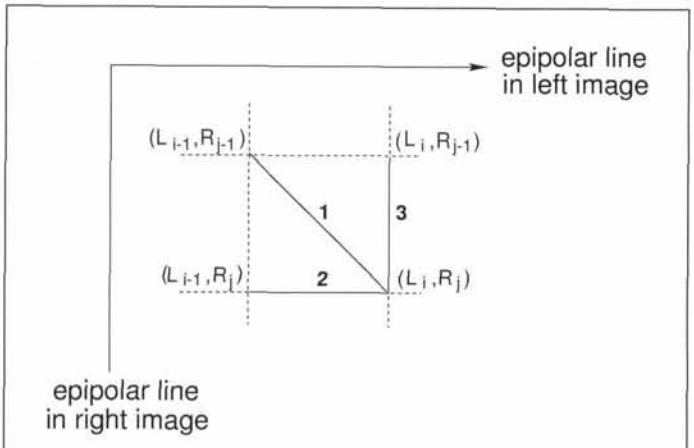


Figure 4. Elementary pathes for dynamic programming. The elementary path (1) denotes a match between the two points L_i and R_j the elementary paths (2) and (3) describe the occlusion of L_i in the right image and the occlusion of R_j in the left image, respectively.

The difficulty of dynamic programming stems from choosing elementary cost functions and the need to preserve interline consistency. In the next two sections, our method for edge matching and area-based matching are detailed, paying special attention to the choice of cost functions and interline consistency.

Edge Point Matching

Dynamic programming has often been used for edge matching, with various methods to achieve interline consistency. Ohta and Kanade (1985) have proposed a simultaneous optimization for all epipolar line pairs, by involving a three-dimensional

search graph. The continuity constraint across lines can also be imposed by post-processing, which requires much less computation (Baker and Binford, 1981; Lloyd *et al.*, 1987). Another solution consists in involving the inter-line consistency in the matching cost function. Wu and Maître (1989) have proposed a consistency cost depending on the disparity differences between two neighboring lines; however, this is subject to error propagation.

We have defined a new approach to match edge points (edgels) which takes the interline consistency into account in the matching cost function. Error propagation is avoided by working in several steps. The originality of the method is to use the distribution of "potential" matches to compute the matching costs.

First, a local elementary similarity cost is computed for each edgel pair. This cost defines potential edgel pairs and potential disparity values. An analysis of the distribution of potential disparity values along edge chains leads to the computation of the final matching cost function. The matching is then achieved by dynamic programming for each epipolar line pair. Finally, an *a posteriori* correction is performed to guarantee the consistency of the 3D chains. The algorithm is detailed in the next subsections.

Edge Points and Attributes

Edges are extracted from each image with the Canny-Deriche filter (Deriche, 1987) followed by hysteresis thresholding. Three quantitative attributes are associated with each edge point (or edgel):

- θ, the local orientation of the intensity gradient, computed modulo 2π, given as output of the Canny-Deriche filter; and
- I_l and I_r, the intensity values on each side of the edge point (left side and right side), along the epipolar line, computed by the "Toboggan" enhancement technique proposed in Fairfield (1990).

Potential Edge Pairs

The *elementary matching cost* between two edge points L_i (left image) and R_j (right image) is defined by

$$match_0(i,j) = \alpha \frac{\min[\Delta I_l(i,j), \Delta I_r(i,j)]}{\Delta I_0} + \beta \frac{\Delta\theta}{\Delta\theta_0}$$

where

$\Delta I_l(i,j) = |I_l(L_i) - I_l(R_j)|$;
$\Delta I_r(i,j) = |I_r(L_i) - I_r(R_j)|$;
$\Delta\theta = |\theta(L_i) - \theta(R_j)$, computed modulo 2π;
$\Delta I_0 = 255$, $\Delta\theta_0 = \pi$ are normalization constants; and
α and β are weight values satisfying $\alpha + \beta = 1$; They are automatically computed for each epipolar line pair, as respectively proportionnal to the intra-line variance of the intensity and the intra-line variance of the gradient orientation.

A pair of edgels (L_i, R_j) is a potential edgel pair if

$$match_0(i,j) < C_{\max}$$

where $C_{\max}$ is the upper bound value for the elementary cost, above which pairs are rejected. Its numerical value expresses a compromise between match density and quality. The disparity value related to a potential edgel pair is called a *potential disparity value.*

Final Matching Cost

A polygonal approximation of the edge chains is performed in each image in order to partition the chains into linear segments. The set **P** of potential edgel pairs is then partitioned in two independent ways: a partition $\mathbf{P}_L$ is defined relative to the segments from the left image, and a partition $\mathbf{P}_R$ is defined relative to the segments from the right image.

More precisely, two potential edgel pairs (L_1, R_1) and (L_2, R_2) belong to the same element of $\mathbf{P}_L$ if L_1 and L_2 belong to the same segment S_L of the left image. These pairs are said to be *consistent* with each other if the corresponding disparity values $d_1 = disp(L_1, R_1)$ and $d_2 = disp(L_2, R_2)$ satisfy the following property:

$$\forall d \in N,\ d_1 \le d \le d_2,\ \exists (L, R) \in \mathbf{P}_L / L \in S_L, \text{ and } d = disp(L, R).$$

The consistency relation between edgel pairs of $\mathbf{P}_L$ is illustrated in Figure 5. It defines an equivalence relation over **P** respecting the partition $\mathbf{P}_L$. A similar consistency relation is defined respecting the partition $\mathbf{P}_R$ issued from the right image.

Let (L_i, R_j) be a potential edgel pair, with L_i belonging to a segment S_L of the left image and R_j belonging to a segment S_R of the right image. If $\ell(S_L) \ge \ell(S_R)$ (ℓ being the segment length), the final matching cost $match(i,j)$ between L_i^c and R_j^c is defined by

$$match(i,j) = \frac{1}{\ell(S_L)} \sum_{L_k^c \in S_L} cmatch_{(i,j)}(k)$$

where $cmatch_{(i,j)}(k)$ is given by

$$cmatch_{(i,j)}(k) = \begin{cases} C_{\max} & \text{if } \mathbf{S}_{i,j,k} = \emptyset \\ Min_l\{match_0(k,l)/R_l \in \mathbf{S}_{i,j,k}\} & \text{otherwise} \end{cases}$$

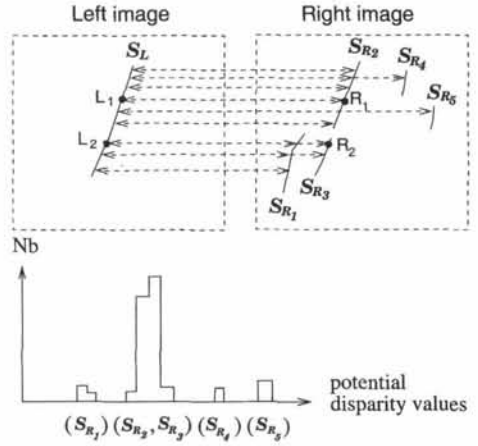


Figure 5. Definition of consistent pairs. Top: Potential pairs related to the segment S_L (potential correspondences are represented by the symbol →). Bottom: Distribution of the potential disparity values related to S_L. The four connected components of the histogram define four sets of consistent pairs. For example the edgel pairs (L_1, R_1) and (L_2, R_2) are consistent. Consistent edgel pairs can correspond to several segments in one of the images, as long as their disparity values belong to the same component of the histogram (for instance, the two line segments S_{R_2} and S_{R_3} define the same component).

and

$$\mathbf{S}_{i,j,k} = \{R_l/(L_k, R_l) \text{ consistent with } (L_i, R_j)\}.$$

The matching cost $match(i,j)$ is averaged over the segment S_L; a contributing cost can be either an elementary cost or the maximal cost $C_{\max}$, according to the content of the set $\mathbf{S}_{i,j,k}$. If $\ell(S_L) \geq \ell(S_R)$, the matching cost is defined in a symmetric way according to the segment S_R. If (L_i, R_j) is not a potential edge pair, then

$$match(i,j) = C_{\max}.$$

According to this definition, $match(i,j) \leq C_{\max}\ \forall(i,j)$. The value $match(i,j)$ decreases with the number of "neighboring" potential matches consistent with (L_i, R_j). Therefore, the interline consistency is implicitely taken into account through the *intra-chain consistency.*

Optimization

Once the matching cost function has been computed for each edgel pair, a correspondence graph is created for each epipolar line pair, consiting of the edgels from the left and right images. In case of adjacent edge points on the epipolar line, only the first and the last are kept. Two additional points L_0 and R_0 define the starting position of the path. The ending location of the path is free within the disparity range.

The matching cost of L_i and R_j is $match(i,j)$. The occlusion cost occ is related to an unmatched edge (occlusion or non-detected corresponding edge point). It is constant and associated with the maximal matching cost by the relation

$$occ = \frac{C_{\max}}{2}.$$

This relation expresses the fact that matching L_i to R_j is "better" than both occlusions of L_i and R_j if and only if $match(i,j) < 2occ$ (see Figure 4).

After the optimal path in the correspondence graph has been computed by dynamic programming, a postprocessing is performed to remove local errors and complete the matching on chains where pixels have not been matched. It is based on the continuity and linearity criteria proposed in Mohan *et al.* (1989): i.e., disparity values are continuous and linear along linear edge segments. Therefore, the disparity values are approximated according to a linear model defined for each segment of the left image, using least-squares estimation.

Area-Based Matching between Edge Points

Principle

In order to produce dense elevation data, area-based matching is performed between the matched edge points. If (L_0^c, R_0^c) and (L_n^c, R_m^c) are two consecutive pairs of edge points on an epipolar line pair, then this stage aims at matching the intervals $]L_0^c..L_n^c[$ and $]R_0^c..R_m^c[$. At this stage, the matching is not limited to edge points, i.e., it concerns any pixel.

Dynamic programming is used again. A correspondence graph is defined for each interval pair. The end points of the path are given by (L_0^c, R_0^c) and (L_n^c, R_m^c). The matching cost between two pixels L_i and R_j is given by

$$match(i,j) = 1 - S^N(i,j)$$

where $S^N(i,j)$ is the centered normalized cross-correlation coefficient computed over a window of size N. Thus,

$$match(i,j) \in [0..2] \forall (i,j).$$

The size of the window is constant for the robustness of the process. The use of adaptive correlation windows depending on intensity edges or matched edges makes the process very dependent on the quality of the edge detection or the edge matching.

As before, the occlusion cost is fixed and related to a minimal threshold $S_{\min}$ on the correlation values: i.e.,

$$occ = \frac{1 - S_{\min}}{2}.$$

Every occlusion is related to a disparity jump. Therefore, when occ decreases, $S_{\min}$ becomes larger and the similarity constraint is stronger, but height discontinuities are more frequent. The numerical value of occ expresses a trade off between the similarity constraint on matched pixels and a smoothness constraint on the surface.

In order to take advantage of this duality, we propose a two-step procedure. A first *radiometrically constrained* matching step with a *low occlusion cost* provides high quality matched pairs without a strong constraint on the surface smoothness, which is very important in urban scenes. A second matching step is then performed to complete the elevation data, which is, on the contrary, *geometrically constrained* by a *high occlusion cost* and based on the reliable matches issued from the first step.

This hierarchical strategy relies on the assumption that local extrema of depth along epipolar lines are recovered as reliable pairs during the first step. This is practically true because local depth extrema are often rather contrasted in the images. Under this assumption, the smoothness constraint of the second step is applied to complete the matching on unmatched areas. In this way, the reliable matches, i.e., those characterized by high similarity, are used to aid in the more ambiguous decisions. The next two subsections describe each of those steps.

Radiometrically Constrained Matching

In this first step, the occlusion cost is quite low (practically, $occ = 0.25$). The input intervals are defined by the matched edges from the previous stage. Using the matched edges as anchor points decreases the number of potential matches, and thus reduces both erroneous matching and computation time. Additionally, the edgel pairs must be verified by the results of the area-based matching: the disparity value of an edgel pair must be close to the disparity values of the neighboring pixels (at least on one side). If an edgel pair is not verified, then it is removed from the matching set. A new interval pair is computed by merging the two intervals separated by the invalid edgel pair, and it is matched again.

Figure 6 shows an example of a matching path followed by the algorithm.

Despite of the low occlusion cost, the disparity map produced by this matching step contains a few errors in homogeneous areas. We propose three postprocessing steps to detect and remove them:

- Shadows are very common homogeneous areas in urban imagery, and they often induce errors in the disparity map. Therefore, shadows are automatically detected in each image. First, the histogram of intensity values is analyzed in order to determine the first local minimum. The corresponding intensity value is used as a threshold for the intensity image. The binary thresholded image is finally filtered by morphological operators. The pairs of pixels corresponding to shadow pixels in each image are removed from the matching set.
- The normalized centered cross-correlation coefficient is invariant to affine transformation on intensity, which is important for the robustness to change in radiometry. But this property sometimes induces errors. In order to detect these errors, we assume that the average radiometric difference between corresponding pixels is locally planar. The image is divided into

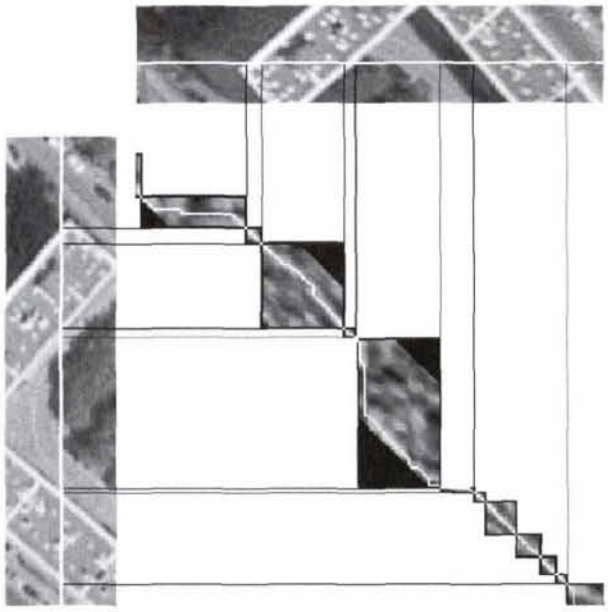


Figure 6. Example of a matching path automatically followed by dynamic programming. The matching path (in white) is superimposed with the cross-correlation scores computed with a window size of 7 by 7 (bright areas correspond to high correlation scores). The path is constrained by the matched edges (the grey areas are not allowed) and by the disparity range (the black areas are not allowed).

rectangular patches, and the plane of radiometric differences is estimated for each patch by least-squares optimization. Outlyers are removed from the matching set.

- The last postprocessing step is justified by the implicit morphology of the surface: the very small regions (an area of a few pixel) isolated in 3D are assumed to be matching errors. They are detected and removed by morphological filtering.

Geometrically Constrained Matching

The radiometrically constrained area-based matching step provides reliable pairs but some disparity values are missing, mainly in homogeneous areas. The last matching step aims at completing elevation data for any area which is not occluded. It is achieved by using a weak radiometric constraint but a strong geometric constraint.

The input intervals are the unmatched intervals from the radiometrically constrained matching step. The occlusion cost is now high (practically, $occ = 0.5$), which defines a very small minimal correlation threshold. It tends to produce a disparity function which varies in a monotone way between two pairs along epipolar lines, in order to minimize the number of occlusions. A geometric smoothness of missing data is actually implicitely assumed.

The information extracted at the previous stage about shadows and radiometric differences is also exploited at this stage. When matching an interval pair whose end point elevations are h_1 and h_2, $h_1 \leq h_2$, then any pair of shadow pixels located within the interval pair and corresponding to the elevation h_1 (the smallest) is supported by a very low matching cost. This expresses that shadow areas are preferably located at a low elevation. Similarly, the putative pixel pairs with a difference in radiometry distant from the plane of differences computed at the previous step are affected by a high matching cost.

As a last step, a morphological filtering along columns is applied in order to remove narrow noise regions parallel to epipolar lines. This kind of noise is characteristic of matching methods based on dynamic programming, when each epipolar line pair is processed independently from the other.

Experimentation and Discussion

The algorithm has been tested on different kinds of images, with various pixel sizes (between 16 and 103 cm on the ground) and various scene contents (Baillard, 1997). The same parameter values have been used for all experiments. An example of an aerial stereo pair is shown in Figure 1, and the main results of the matching process are illustrated in Figure 7. Another example of a disparity map automatically computed is shown on Figure 8.

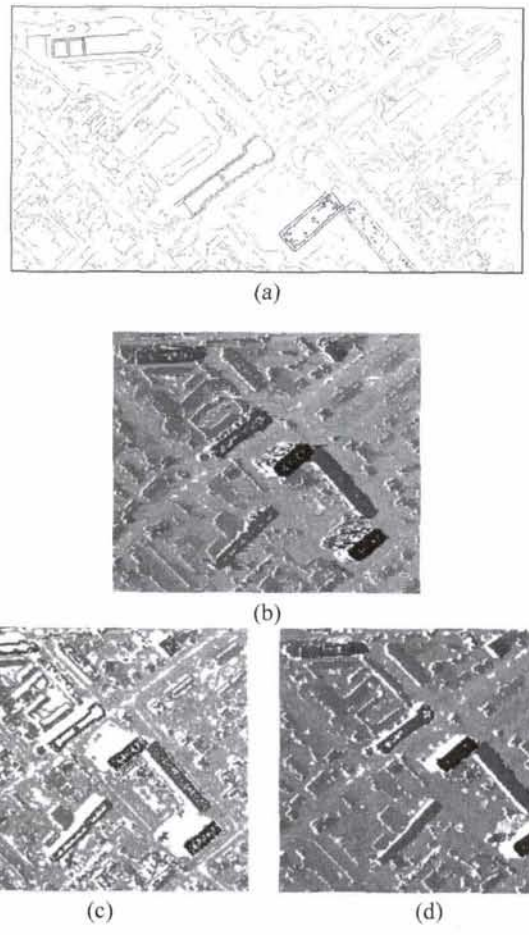


Figure 7. Disparity maps produced for the pair of Figure 1, with and without the hierarchical approach: the darkest grey levels correspond to the highest elevations, and the white areas to unmatched pixels. (a) Disparity map after edge point matching (subpart). (b) Disparity map produced with a one-step dynamic programming process ($occ = 0.5$). (c) Disparity map after the radiometrically constrained area-based matching step ($occ = 0.25$) (d) Disparity map after the geometrically constrained area-based matching step ($occ = 0.5$). In comparison to (b), the disparities are much less noisy in the neighborhood of large differences in height. The hierarchical strategy is of prime importance for the quality of the results.

(a)

(b)

Figure 8. (a) Left intensity image (1105 by 1024 size, pixel 40 cm, disparity range 70 pixels); (b) Final disparity map.

The edge matching step provides reliable 3D information about most edges of the image. Depending on the input images, between 60 and 80 percent of the edge points are matched. Lost edges often correspond to vegetation areas, which do not verify the linearity hypothesis.

After the first radiometrically constrained area-based matching step, more than half of the pixels have been matched. Unmatched pixels mostly correspond to hidden area (occlusions), vegetation, and homogeneous areas (low correlation scores). The disparity values are reliable, and local extrema have been located in 3D.

The final disparity maps are dense with very little noise, even when a large disparity range is used. The height discontinuities have been preserved, although slightly delocalized. This is due to the use of a fixed-size correlation window, which produces a biaised similarity measure in the neighborhood of discontinuities. However, the global shape of the objects has been preserved. Experiments have shown than a correlation window limited by intensity edges is too sensitive to the quality of edge detection. It leads to irregularly shaped objects when intensity edges are missing or broken (very frequent in complex urban images at a resolution of over 40 cm per pixel), and to noisy elevations over textured areas like vegetation. The disparity accuracy is about 1 pixel, and the planimetric accuracy is smaller than 3 to 4 pixels (half the size of the correlation window).

The quality of the result is partly due to the dynamic programming optimization process, which relies highly on consistency constraints. More especially, the duality between *discontinuities and occlusions is a key point for successful* matching in urban environments.

The hierarchical strategy is also essential. Figure 7b shows a disparity map produced by a one-step dynamic programming process. When compared to Figure 7d, it shows that, given a similar matching density, errors are significantly reduced by involving a hierachical process. This is especially true in the neighborhood of large occlusions. In addition, the total processing time has decreased by half. The efficiency of the hierarchical strategy stems from a relevant reduction of the search space: edges, well-contrasted pixels, and homogeneous areas are each processed in an appropriate way. The radiometric information provided by reliable points (shadows and radiometric differences) makes the final elevation data even more robust.

The characteristics of these results allow interesting perspectives for DEM production in urban scenes.

Application to DEM Production

The algorithm has been used experimentally for the automatic production of urban DEMs on a large scale. For that purpose, it has been computationnally adapted to the processing of large sets of data (big input images). The control of the search interval for corresponding points is an important aspect, both to limit false matches and to keep the processing time small. The parametrization of the disparity range is therefore performed using a 25-m-grid DTM available for all of France (accuracy 5m.). This external altimetric database provides elevations of the ground surface and, therefore, a local lower bound value for height. This information dynamically constrains the search interval to be explored during the matching process. The highest elevations can be automatically computed by matching subsampled images, or provided by a coarse hand-made map of highest elevations (which can be expressed relatively to the lowest ones). This latter solution has been chosen for our experiments, because it is quick and reliable. In addition, all bodies of water have been excluded from the correlation area because of specular reflections occuring at these places.

An example of a 5000- by 10000-pixel stereo pair is shown in Figure 9 (subregion of Paris, resolution 40 cm per pixel, sampled in an epipolar geometry). The DEM computed with the method is shown in Figure 10. The interactive preprocessing step lasted 30 minutes, consisting of making the map of highest elevations and delineating borders of the Seine River. The automatic matching of the whole area was achieved in 5 hours on an ordinary workstation. This shows that the method can be used in practical situations for the production of urban DEMs on a large scale.

These kinds of DEMs can also be used in the production of 3D perspective views of a scene. Figure 11 shows a 3D view of Paris, automatically computed from the DEM of Figure 10, and texture mapped using the two input images. For that purpose, the DEM has been segmented into ground and above ground areas with the method described in (Baillard and Maitre,

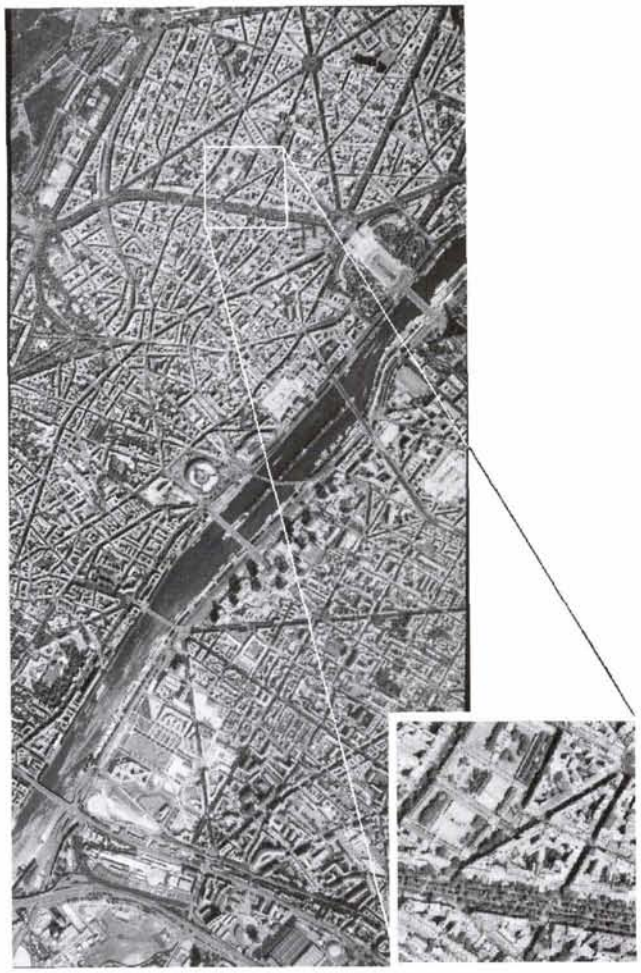


Figure 9. Example of a large aerial image of Paris, and a subregion of it. The large image is about 5,000 by 10,000 pixels, at a resolution of 40 cm on the ground.

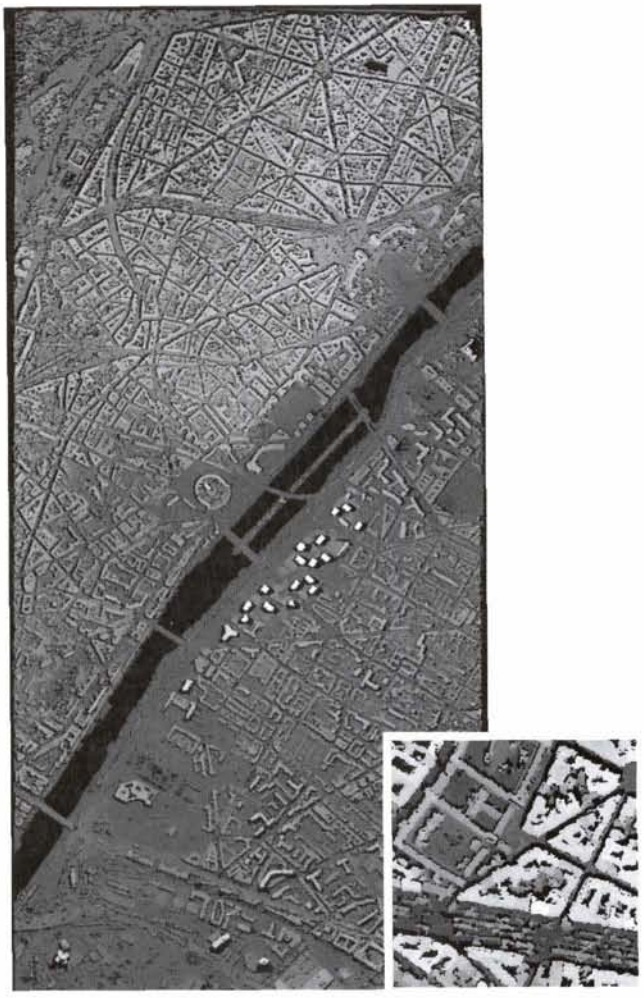


Figure 10. DEM automatically produced by stereo matching for the example of Figure 9 (projected onto a cartographic reference). The unmatched areas of the DEM (in black) are due to hidden areas (occlusions). Water areas have been removed because of specular reflection.

1999). A DTM (digital terrain model) of the scene was simultaneously derived from the detected ground regions. The DTM was used to complete 3D data for the occluded areas that were not visible on both images. The textural data came from an "ortho-image" of the scene (a radiometric image where all the projective distortion has been corrected). The ortho-image was derived from the DEM and both radiometric images, and corrected by the planar model of the grey-level differences described in the section on Radiometrically Constrained Matching. The occluded areas that were visible on one image only were mapped by the visible intensity. As for areas occluded in both images (lateral occlusions), the intensity could be interpolated without visual failures in the case of small surfaces.

The perspective view which was produced shows a few defaults but it is visually quite good, and it can be used for applications such as urban planning, impact study, or advertisement.

Conclusion

We have presented a matching algorithm dedicated to urban stereo pairs. It consists of successive steps based on dynamic programming. The originality of the method is two-fold: the explicit use of duality between discontinuities and occlusions, and the definition of a relevant hierarchical strategy.

Experimentation has shown that this matching algorithm has many advantages. It provides dense elevation values (only occluded points are not matched), and preserves height discontinuities. The 3D data are reliable, even in the neighborhood of large differences in height or in homogeneous areas. In addition, the algorithm is fast, it is robust to the variability of images (scene content and pixel resolution), and it does not require a fine tuning of parameters. Its main drawback is the relative planimetric inaccuracy in the neighborhood of height discontinuities, which has been conceded to achieve robustness. For many practical applications, robustness and density are more important than high accuracy. However, for applications where a very good planimetric accuracy is required, the same method could be applied with an adaptive correlation window (Cord *et al.*, 1998).

These characteristics make the matching algorithm particularly interesting for the production of urban DEMs. It has been successfully implemented within an industrial context to produce urban DEMs on a large scale. It also provides a good basis for the production of ortho-images or 3D perspective views. Within a fine building reconstruction process, the provided DEM can be used to detect regions of interest (buildings for example)

Figure 11. Two 3D perspective views of the 3D model produced from the DEM shown in Figure 10 and texture mapped from the original images.

where a geometric refinement is expected (Baillard and Maitre, 1999). The local and contextual information related to regions of interest can help for an accurate reconstruction in many ways: e.g., selection of relevant features, choice of an appropriate reconstruction model, and automatic parametrization. Therefore, the matching algorithm presented in this paper can be very useful both for the production of DEMs on a large scale and as a starting point for a coarse-to-fine process dedicated to an accurate and complete 3D reconstruction.

Acknowledgment

Financial support was provided by the IGN. The authors are very grateful to Pr Henri Maître, from ERNST Paris, for his helpful advice and for supervising the project.

References

Baillard, C., 1997. *Analyse d'images aériennes stéréoscopiques pour la restitution 3-D des milieux urbains*, PhD Thesis, ENST 97-E-018 (available at http://www-isis.enst.fr/Kiosque/theses/MANUSCRITS/) (in French).

Baillard, C., and H. Maitre, 1999. 3D reconstruction of urban scenes from aerial stereo imagery: A focusing strategy, *Computer Vision and Image Understanding*, 76(3):244–258.

Baillard, C., C. Schmid, A. Zisserman, and A. Fitzgibbon, 1999. Automatic line matching and 3-D reconstruction of buildings from multiple views, *Automatic Extraction of GIS Objects from Digital Imagery, International Archive of Photogrammetry and Remote Sensing*, 32(Part3-2W5):69–80.

Baker, H.H., and T.O. Binford, 1981. Depth from edge and intensity based stereo, *Proc. 7th Int. Joint Conf. on Artificial Intelligence*, August, Vancouver, B.C., Canada, pp. 631–636.

Belhumeur, P.N., and D. Mumford, 1992. A Bayesian treatment of the stereo correspondence problem using half-occluded regions, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 15–18 June, Champaign, Illinois, pp. 506–512.

Berthod, M., L. Gabet, G. Giraudon, and J.L. Lotti, 1995. High-resolution stereo for the detection of buildings, *Automatic Extraction of Man-Made Objects from Aerial and Space Images* (A. Grün, O. Kübler, and P. Agouris, editors), Birkhauser Verlag, Basel, Switzerland, pp. 135–144.

Bignone, F., O. Henricsson, P. Fua, and M. Stricker, 1996. Automatic extraction of generic house roofs from high resolution aerial imagery, *Proc. of 4th European Conf. on Computer Vision*, 16–18 April, Cambridge, England, pp. 85–96.

Birchfield, S., and C. Tomasi, 1998. Depth discontinuities by pixel-to-pixel stereo, *Proc. of 6th Int. Conf. on Computer Vision*, 03–08 January, Bombay, India, pp. 1073–1080.

Cochran, S.D., and G. Medioni, 1992. 3-D surface description from binocular stereo, *IEEE Trans. on PAMI*, 14(10):981–994.

Cord, M., N. Paparoditis, and M. Jordan, 1998. Dense, reliable, and depth discontinuity preserving DEM computation from very high resolution urban stereopairs, *ISPRS Symp.*, 13–17 July, Cambridge, England, pp. 49–56.

Dang, T., O. Jamet, and H. Maitre, 1994. Applying perceptual grouping and surface models to the detection and stereo reconstruction of building in aerial imagery, *Int. Archives of Photogrammetry and Remote Sensing*, 30:165–172.

Deriche, R., 1987. Using Canny's Criteria to derive a recursively implemented optimal edge detector, *Int. J. of Computer Vision*, 1:167–187.

Dhond, U.R., and J.K. Aggarwal, 1995. Stereo matching in the presence of narrow occluding objects using dynamic disparity search, *IEEE Trans. on PAMI*, 17(7):719–724.

Fairfield, J., 1990. Toboggan contrast enhancement for contrast segmentation, *Proc. of 10th Int. Conf. on Pattern Recognition*, 01 June, Atlantic City, New Jersey, pp. 712–716.

Fradkin, M., M. Roux, H. Maitre, and U.M. Leloglu, 1999. Surface reconstruction from multiple aerial images in dense urban areas, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 23–25 June, Fort Collins, Colorado, pp. 262–267.

Fua, P., 1991. Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities, *Proceedings, Int. Joint Conf. on Artificial Intelligence*, 24–30 August, Sydney, Australia.

Geiger, D., B. Ladendorf, and A.L. Yuille, 1995. Occlusions and binocular stereo, *Int. j. of Computer Vision*, 14:211–226.

Girard, S., P. Guerin, H. Maitre, and M. Roux, 1998. Building detection from high resolution colour images, Proceedings, *SPIE Europto Image and Signal Processing for Remote Sensing IV (S. Serpico, editor)*, September, Barcelona, Spain, 3500:278–289.

Haala, N., and M. Hahn, 1995. Data fusion for the detection and reconstruction of buildings, *Automatic Extraction of Man-Made Objects from Aerial and Space Images* (A. Grün, O. Kübler, and P. Agouris, editors), Birkhauser, Basel, Switzerland, pp. 211–220.

Hannah, M.J., 1989. A system for stereo image matching. *Photogrammetric Engineering & Remote Sensing*, 55(12):1765–1770.

Henricsson, O., 1998. The role of color attributes and similarity grouping in 3-D building reconstruction, *Computer Vision and Image Understanding*, 72(2):163–184.

Hoff, W., and N. Ahuja, 1989. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection, *IEEE Trans. on PAMI*, 11(2):121–136.

Horaud, R., and T. Skordas, 1989. Stereo correspondence through feature grouping and maximal cliques, *IEEE Trans. on PAMI*, 11(11):1168–1180.

Kanade, T., and M. Okutomi, 1994. A stereo matching algorithm with an adaptative window: Theory and experiment, *IEEE Trans. on PAMI*, 16(9):920–932.

Leloglu, U., M. Roux, and H. Maitre, 1998. Dense urban DEM with three or more high-resolution aerial images, Proceeding, *ISPRS Symposium on GIS - Between Visions and Applications, September, Stuttgart, Germany*, pp. 347–352.

Lloyd, S.A., E.R. Haddow, and J.F. Boyce, 1987. A parallel binocular stereo algorithm utilizing dynamic programming and relaxation labelling, *Computer Vision, Graphics and Image Processing*, 39:202–225.

Lotti, J.-L., and G. Giraudon, 1994. Correlation algorithm with adaptative window for aerial image in stereo vison, *SPIE*, 2315:76–87.

Luo, A., and H. Burkhardt, 1995. An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions, *Int. J. of Computer Vision*, 15:171–188.

Maitre, H., and W. Luo, 1992. Using models to improve stereo reconstruction, *IEEE Trans. on PAMI*, 14(2):269–277.

Mohan, R., G. Medioni, and R. Nevatia, 1989. Stereo error detection, correction, and evaluation, *IEEE Trans. on PAMI*, 11(2):113–120.

Moons, T., D. Frere, J. Vandekerckhove, and L. Van Gool, 1998. Automatic modelling and 3D reconstruction of urban house roofs from high resolution aerial imagery, *Proc. of 5th European Conf. on Computer Vision*, 02–06 June, Freiburg, Germany, pp. 410–425.

Noronha, S., and R. Nevatia, 1997. Detection and description of buildings from multiple aerial images, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 17–19 June, Puerto Rico, pp. 588–594.

Ohta, Y., and T. Kanade, 1985. Stereo by intra- and inter-scanline search using dynamic programming, *IEEE Trans. on PAMI*, 7(2):139–154.

O'Neill, M.A., and M.I. Denos, 1992. Practical approach to the stereo matching of urban imagery, *Image and Vision Computing*, 10(2):89–98.

Otto, G.P., and T.K.W. Chau, 1989. Region-growing algorithm for matching of terrain images, *Image and Vision Computing*, 7(2):83–94.

Paparoditis, N., M. Cord, M. Jordan, and J.-P. Cocquerez, 1998. Building detection and reconstruction from mid- and high-resolution aerial imagery, *Computer Vision and Image Understanding*, 72(2):122–142.

Roux, M., and D.M. McKeown. 1994. Feature matching for building extraction from multiple view, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 21–23 June, Seattle, Washington, pp. 46–53.

Schmid, C., and A. Zisserman, 1997. Automatic line matching across views, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 17–19 June, Puerto Rico, pp. 666–671.

Weidner, U., and W. Foerstner, 1995. Towards automatic building extraction from high-resolution digital elevation models, *ISPRS J. of Photogrammetry and Remote Sensing*, 50(4):38–49.

Wu, Y., and H. Maitre, 1989. Mise en correspondance d'images stéréo par programmation dynamique untilisant la coherence inter lignes, *12th colloque GRETSI*, Juan-les-Pins, June France, pp. 751–754 (in French).

(Received 22 July 1998; accepted 25 June 1999; revised 25 October 1999)